Estimating the Euro Area output gap through a Non-stationary DFM and a BVAR model

Dario Alfredo De Falco*
Francesco Sbrana†
Filippo Palandri‡
Giovanni Brendolin§

October 2023

Abstract

We measure the output gap of the Euro area, using a limited dataset composed of both macroeconomic and financial variables, by means of two different methods: the first based on a dynamic factor model, the second based on the Beveridge-Nelson decomposition applied to a BVAR. Our results show that the model developed by Morley et al. [2023] seems to be more reliable in a setting in which the available variables are few, with the estimated output gap that is in line with the CEPR chronology of the business cycle for the euro area. On the other hand, the Barigozzi and Luciani [2023] model's fit seems to improve as the number of series used increases. Moreover, our findings confirm the importance of including financial variables into the output gap estimations, as well as the significant role played by hours worked, differently from US potential output estimations where the role of hours worked is replaced by unemployment.

1 Introduction

One of the goal of the stabilization policy is to reduce the output gap, broadly defined as the deviation of the observed GDP from its trend, the potential output, during economic downturns. Potential output is a measure of the economy's ability to generate output and it is a crucial input for long-term projection, while the output gap is important for the formulation of monetary policy, since it measures the degree of slack within the aggregate economy (Morley et al. [2023]). The output gap is probably the most complete concept to describe the cyclical position of an economy Graff and Sturm [2012]; however, both output gap and potential output are latent variables that need to be estimated, and there is neither a unique definition of potential output nor a sole academic approach on how to estimate it. Hence, due to this measurement uncertainty, strong criticism has been directed to the importance of the output gap as a macroeconomic indicator. For instance, according to Orphanides [2003], the underestimation of the output gap was responsible for the US's high inflation rate in the 1970s.

One of the earlier works on measuring potential output was the one by Okun [1962], which led to the development of the Okun's law, an empirical relationship between output and unemployment. From that year, a wide literature start arise in the measurement of potential output in various countries. As an example, Orphanides and Norden [2002] shows that conventional statistics measures produce unreliable estimates of the US output gap in real time,

*Dario Alfredo De Falco worked on the coding part on MATLAB. Moreover, he took care of section 4.1, 4.4 and 4.5.

†Francesco Sbrana collected all the data used and took care of their description in section 3. Moreover, he took care of section 4.2, 4.3 and 4.5.

‡Filippo Palandri took care of the Methodology part, with specific reference to the exposition of the Morley et al. [2023] model (section 2.3, 2.4 and 2.5). He worked with Francesco Sbrana for section 4.3.

§Giovanni Brendolin took care of the Methodology part, with specific reference to the exposition of the Barigozzi and Luciani [2023] model (section 2.1 and 2.2). He worked with Dario De Falco for section 4.1.

because they are subject to large revision and to unreliability of end-of-sample estimates; Marcellino and Musso [2011] find that estimates of the output gap in the euro area are also highly uncertain, attributing this finding to parameter instability and model uncertainty instead of data revision.

As said, one of the central issue in measuring the output gap is that there is no single theoretical definition of the potential output. Following Kiley [2013], we describe the three main approaches to estimate the output gap, each of which simply derives from a different definition of the potential output. It has to be noticed that any of these definition yields different conclusions in the estimation. According to the statistical approach Beveridge and Nelson [1981], potential output is the long-run stochastic trend of output. This definition allows to measure the output gap without relying too heavily on theory, but relying only on some theoretical assumptions Basu and Fernald [2009], using filters that separates the trend from cyclical or low-frequency fluctuations. The production-function approach instead, defines the potential as the level of output consistent with all resources employed at their full potential, and this is the definition used by the Congressional Budget Office in their measures. Lastly, according to the New Keynesian approach, potential output is the level of real GDP reached in absence of nominal and financial frictions and of inefficient shocks.

In this elaborate, we try to estimate the Euro area output gap using two different methods. The first one have been developed by Barigozzi and Luciani [2023] and is closely related to the statistical approach. The model works by estimating a large-dimensional non-stationary dynamic factor model isolating the factors driving the common dynamics, and then decomposing the estimated common factors into common trends and common cycles using principal component analysis on the estimated factors. The second model we use has been developed by Morley and Wong [2020] and it applies the Beveridge and Nelson [1981] decomposition based on a large Bayesian vector autoregression (BVAR) model, but accounting for the fact that the euro area is an open economy through block-exogeneity assumption, as in Morley et al. [2023]. The two models are compared with each other and put against a few benchmark measures, those estimated by international institutions, such that the IMF, the European Commission and the OECD, and those measures obtained by using more widely spread methods such as the UCM (Unobserved Component Model) and the HP filter method. Evaluating these methods, however, is a difficult task. Unlike standard out-of-sample forecast experiments with observed data, potential output is latent, so there is no "truth" against which compare the results.

Our estimations are based on a relatively limited set of variables from the Euro area, composed of both macroeconomic and financial time series, avoiding to address the Covid19 pandemic period. A few variables related to the international economy are also included. While the small number of variables seems to work quite good in predicting the output gap with the Morley and Wong [2020] model, we expect this can cause estimation problems while using the Barigozzi and Luciani [2023] method. Indeed, in order to isolate the common factors driving the common dynamics, using a large number of variables is crucial, since aggregating a large number of variables allows us to disentangle macroeconomic fluctuations from idiosyncratic dynamics (Stock and Watson [2002]; Bai and Ng [2002]). We thus expect the BVAR model to be more reliable in this setting.

Our estimations show that the model developed by Morley et al. [2023] seems to be more reliable in a setting in which the available variables are few, while the Barigozzi and Luciani [2023] model's fit diverges substantially from other institutional estimates. However, the estimation seems to improve as the number of series used increases accordingly, suggesting that perhaps this model is more adapt for a large number of explanatory variables, as noted in the literature. Moreover, our findings confirm the importance of including financial variables into the output gap estimations, as well as the significant role played by hours worked. Indeed, the latter brings with them a big amount of information content, resulting to be one of the most important variable in the output gap estimation for the Euro area, differently from US output gap estimations where this role has been played, instead, by unemployment.

The paper is structured as follows: section 2 describes the methodology for both the two models, addressing also the theoretical assumptions behind them; section 3 describes the dataset used, the transformations made and how we decided to address the foreign block; section 4 reports our results and finally section 5 contains our conclusions.

2 Methodology

2.1 Barigozzi-Luciani model specification

Different definitions of output gap lead to different estimation methods. The New Keynesian approach defines it as simply the deviation of the actual (observed) output from the potential (unobserved) one (Justiniano et al. [2013]). The latter is the output that would be reached if all financial and/or nominal frictions and inefficiencies did not exist. The production-function approach (from the Congressional Budget Office CBO [2001]) defines potential output as the one stemming from current state of the art of technology A and efficient use of capital (K) and labor (L).[1] What we focused on is the statistical approach (Beveridge and Nelson [1981]) and partially adopted by Barigozzi and Luciani [2023]. According to this approach the potential output is determined by its long-run stochastic trend. It follows, hence, that our estimation is not based on any theoretical model and is primarily data driven. This is a considerable difference with respect the output gap computed by statistical agencies and public institutions.

The method works in two step:

1. **step one** involves isolating the common dynamics of macroeconomic and financial variables through the estimation of a dynamic factor model; through this step, aggregate dynamics of the economy determined by macro-level shocks are isolated, while leaving the idiosyncratic dynamic and measurement error of each series out.

2. **step two** disentangles the long run trend and the common cycle through principal component analysis on the estimated factor. By doing so, we identify the long-run stochastic trend of output (i.e. its persistent dynamic) and the short-run dynamic (due to temporary and business cycle-led shocks).[2]

Though our data (described in section 3), we can capture real, nominal and financial shocks which drive comovements across the Euro Area (EA).

2.1.1 Step one - DFM

We start from our observations y_{it} of n time series recorded for T periods, hence $\boldsymbol{y_t} = \{y_{it} : i = 1, ..., n\}$ for $t = 1, ..., T$ is a $n \times 1$ vector. It is assumed that the series evolves around a deterministic long-run trend (denoted as $\mathcal{D}_{it}$). This trend captures the non-stationary nature of our variables; indeed, many macroeconomic time series (GDP itself for example) are known to grow over time and hence are not stationary. It is important to point out that, however, this deterministic component need **not** to be common to all variables. Indeed, it is modelled as $\mathcal{D}_{it} = b_{it-1} + \mathcal{D}_{it-1} + \epsilon_{it}$ and b_{it} is a random walk (for η_{it} being its disturbance term). We assume, hence, for some $n_b \in (0, n)$ series the effect of the deterministic component is different from zero and constant over time (i.e. $b_{it} = b_i \neq 0$). For these series the time trend becomes $\mathcal{D}_{it} = a_i + b_i t$. The remaining $n - n_b$ series, instead, are assumed not to be driven by a time trend, i.e. $b_{it} = 0$ and $\mathcal{D}_{it} = a_i$.[3] The only exception to these two cases is represented by GDP. Indeed, following Barigozzi and Luciani [2023], we allow for a time-varying effect of the linear trend on the output variable (i.e. $b_{GDP,t} \neq b_{GDP}$). This allows to capture long-trend dynamics of GDP in advanced countries like "secular stagnation", postulated by Summers [2016].

The deviations from this trend are, by construction, accounted either by (i) some macro-shocks affecting (possibly in an heterogenous fashion, being the loadings i-specific) all series in our dataset, or by (ii) idiosyncratic shocks that are specific to each series and/or measurement errors. (i) are modelled through q (unobserved) dynamic factors

[1] Intuitively, it's the outcome of $Y = Af(K, L)$.

[2] According to the statistical approach for the output gap estimation (Beveridge and Nelson, 1981), indeed, potential output coincides with the long-run stochastic trend of output.

[3] Notice that to derive these expressions for $\mathcal{D}_{it}$ for different series we are implicitly relying on specific distributional assumptions on ϵ_{it} and η_{it}. In particular, for both the n_b and the $n - n_b$ series we are setting $\epsilon_{it} = \eta_{it} = 0$. These assumptions restrict the mean of the variables (but for GDP) to be either a linear trend or a constant.

$\boldsymbol{f}_t = (f_{1t} \dots f_{qt})'$, while (ii) is accounted by the idiosyncratic component ξ_{it}. Hence, the model we estimate is

$$y_{it} = \mathcal{D}_{it} + \boldsymbol{\lambda}_i'(L)\boldsymbol{f}_t + \xi_{it} \tag{1}$$

where L is the lag operator, such that $\boldsymbol{\lambda}_i(L) = \sum_{l=0}^{s} \boldsymbol{\lambda}_{il} L^l$ for l being the number of lags and $s \geq 0$.[4] Notice that, $\forall l$, $\boldsymbol{\lambda}_{il} = (\lambda_{i1l} \dots \lambda_{iql})'$ is a q-dimensional vector; in other words, at each of the s lags, each of the n series is *allowed* to load differently on each of the q factors.

2.1.2 Step two - PCA

Once we've estimated the model in equation (1) (the procedure is described in subsection 2.2), we move on to the principal component analysis (PCA). The aim is to decompose the estimated common factors into (long-run) trend and the cycle (*trend-cycle decomposition*).

We assume that the q factors $\boldsymbol{f}_t$ are cointegrated of order d. Therefore, $\exists \boldsymbol{\beta}$ $(q \times d)$ such that the error correction component (EC) $\boldsymbol{\beta}'\boldsymbol{f}_t \sim I(0)$; in other words, there exist d linear combinations of the q factors which result in a stationary process. It follows, that we can write $\boldsymbol{f}_t$ as:

$$\boldsymbol{f}_t = \boldsymbol{\psi}_1 \boldsymbol{\tau}_t + \boldsymbol{\gamma}_t \tag{2}$$

where $\boldsymbol{\tau}_t$ is a r-dimensional vector of common trends in the factors (i.e. it is the non-stationary component) and $\boldsymbol{\gamma}_t$ is a q-dimensional stationary vector. We know that the cointegration relationship described in equation (2) is not unique and, therefore, we have an identification issue which needs further assumptions to be solved. We assume that the r common trends are linear combinations of the common factors, i.e.

$$\boldsymbol{\tau}_t = \boldsymbol{B}'\boldsymbol{f}_t \tag{3}$$

where $\boldsymbol{B}$ is $q \times r$. By plugging (3) into (2), we get

$$\boldsymbol{f}_t = \boldsymbol{\psi}_1 \boldsymbol{B}'\boldsymbol{f}_t + \boldsymbol{\gamma}_t$$

which yields

$$\boldsymbol{\gamma}_t = (I_q - \boldsymbol{\psi}_1 \boldsymbol{B}')\boldsymbol{f}_t \tag{4}$$

Now, since $(I_q - \boldsymbol{\psi}_1 \boldsymbol{B}')$ is a $q \times q$ and $\boldsymbol{\gamma}_t$ is a stationary vector we can write $(I_q - \boldsymbol{\psi}_1 \boldsymbol{B}') = \boldsymbol{\psi}_2 \boldsymbol{\beta}'$ for some $q \times d$ matrix $\boldsymbol{\psi}_2$. Hence, we rewrite (4) as a linear combination of the EC component

$$\boldsymbol{\gamma}_t = \boldsymbol{\psi}_2 \boldsymbol{\beta}'\boldsymbol{f}_t \tag{5}$$

where $\boldsymbol{\beta}'\boldsymbol{f}_t$ is a d-dimensional vector of stationary processes. In other words, it contains the d stationary common cycles. On the other hand, $\boldsymbol{\tau}_t$ contains the r non-stationary commond trends. So, finally, we can write the factors estimated in step one as composed by a trend and a cycle component.

$$\boldsymbol{f}_t = \boldsymbol{\psi}_1 \boldsymbol{\tau}_t + \boldsymbol{\psi}_2 \boldsymbol{c}_t \tag{6}$$

where $\boldsymbol{c}_t = \boldsymbol{\beta}'\boldsymbol{f}_t$. Intuitively, hence, the common factors in our series evolve according to two drivers:

1. long-run trends $\boldsymbol{\tau}_t$ which determine persistent increases of the mean of the series over time (i.e. captures the non-stationary dynamics); this long-run component represents the potential GDP.

2. stationary fluctuations around these trends $\boldsymbol{c}_t$ which captures the business-cycle deviations of GDP from the

[4] In our estimation, we'll use 1 lag, i.e. the effect of the factor at time t dies out *after* $t + 1$.

potential, i.e. it measures the output gap.

To conclude the subsection, we must say that we're assuming contemporaneous uncorrelation between $\boldsymbol{\tau}_t$ and $\boldsymbol{c}_t$. In other words, the cycle (trend) at time t is not systematically related to the trend (cycle) at time t. This is economically reasonable, indeed from economic theory we can safely assume that that shocks impacting the business cycle (like monetary or fiscal expansions) will have no immediate impact on the trend, which is likely to be the result of past shocks (which have impact *over* time) as well as long term-dynamics. Long-run growth, for example, is not determined by a monetary expansion, which is likely to boost demand in the short-term, but won't impact the predetermined trend that the country is following at the time of the shock. On the other hand, the reverse is also true: determinants of long-run GDP (like demographic factors) are typically slow-moving and hence have a negligible immediate impact on the business cycle. This assumption, however, does not rule out lagged dependencies between the trend and the cycle. Indeed, we are allowing correlation between first differences of $\boldsymbol{\tau}_t$ and $\boldsymbol{c}_t$.[5] It follows that contemporaneous uncorrelation, together with the assumption in equation (3) allows us to get that $\boldsymbol{B} = \boldsymbol{\beta}_\perp$, such that $\boldsymbol{\beta}'_\perp \boldsymbol{\beta} = \mathbf{0}$ $(d \times d)$. The reason for this lies in the fact that if (i) $\boldsymbol{c}_t = \boldsymbol{\beta}' \boldsymbol{f}_t$, (ii) $\boldsymbol{\tau}_t = \boldsymbol{B}' \boldsymbol{f}_t$ and (iii) $\boldsymbol{c}_t$ and $\boldsymbol{\tau}_t$ are orthogonal, then it must be that $\boldsymbol{\beta}' \boldsymbol{f}_t$ and $\boldsymbol{B}' \boldsymbol{f}_t$, and hence $\boldsymbol{\beta}$ and $\boldsymbol{B}$ are orthogonal too. In other words, if we project the common factors onto the space orthogonal to the common cycle component, we are isolating the part unrelated to the common cycle, that is the common trend. Therefore, we can rewrite the common factor as

$$\boldsymbol{f}_t = \boldsymbol{\beta}_\perp \boldsymbol{\tau}_t + \boldsymbol{\beta} \boldsymbol{c}_t \tag{7}$$

so that the original model in (1) becomes

$$y_{it} = \mathcal{D}_{it} + \boldsymbol{\lambda}'_i(L) \boldsymbol{\beta}_\perp \boldsymbol{\tau}_t + \boldsymbol{\lambda}'_i(L) \boldsymbol{\beta} \boldsymbol{c}_t + \xi_{it} \tag{8}$$

where we can see

- *POTENTIAL OUTPUT* given by $\mathcal{D}_{it} + \boldsymbol{\lambda}'_i(L) \boldsymbol{\beta}_\perp \boldsymbol{\tau}_t$, i.e. the deterministic and the stochastic long-run trend;
- *OUTPUT GAP* given by $\boldsymbol{\lambda}'_i(L) \boldsymbol{\beta} \boldsymbol{c}_t$, i.e. the difference between what is observed and the potential.

At this point, hence, we estimate $\boldsymbol{\beta}$ and $\boldsymbol{\beta} \perp$ through PCA.

2.2 Barigozzi-Luciani estimation

First of all, we need the model to be written in state-space form.[6] Hence we define:

1. the *observation equation*, which links the observed variables to unobserved (latent) factors; this is equation (1).
2. the *transition equation*, governing evolution of the state variable (i.e. the factors) over time. To this aim, we assume that it follows a (vector) autoregressive process of order p (VAR(p)):

$$\boldsymbol{f}_t = \boldsymbol{A}(L) \boldsymbol{f}_{t-1} + \boldsymbol{u}_t \tag{9}$$

   where $\boldsymbol{A}(L) = \sum_{l=0}^{p-1} \boldsymbol{A}_l L^l$.

Equations (1) and (9) define the model in state-space form, which can be estimated by quasi-maximum likelihood through expectation maximization (EM) algorithm.[7]
Before starting with estimation, to check for which variable include a linear trend or not, we test for significance of

[5] To be clear,this assumption still allows $E(\Delta\tau_{it}\Delta c_{jt}) = E[(\tau_{it} - \tau_{it-1})(c_{jt} - c_{jt-1})] \neq 0$ due to the fact that $E(\tau_{it} c_{js}) \neq 0$ for $t \neq s$.
[6] This is to provide a framework for modelling dynamics driven by hidden factors.
[7] For the empirical estimation, we run (following the procedure used by Barigozzi and Luciani [2023]) the model for $q = 2, 3, 4$. The results reported in this elaborate are those obtained using $q = 4$. More details in the Appendix.

the average of Δy_{it} (for all variables but for GDP).[8] The procedure starts by estimating the loadings $\boldsymbol{\lambda}_i$ from (1) through PCA on differenced data (to avoid spurious effect due to non-stationary series), getting $\hat{\boldsymbol{\lambda}}_i$. We proceed by estimating the q factors $\boldsymbol{f}_t$ by projecting our series y_{it} onto the estimated loadings $\hat{\boldsymbol{\lambda}}_i$, getting $\hat{\boldsymbol{f}}_t$. We use these estimated factors to run an OLS regression of equation (9) and obtain $\hat{\boldsymbol{A}}(L)$.
With these estimates, we run the first iteration of the EM procedure and then proceed as follows:

- **E-step**: at each iteration, we get new estimates of the the state variable, the deterministic and the idiosyncratic component in (1) and their conditional covariance, through the Kalman filter and the Kalman smoother (they both work recursively but in opposite direction);
- **M-step**: using these estimates, we build the expected log-likelihood; we proceed by maximizing it with respect to $\boldsymbol{\varphi} = (\boldsymbol{\Lambda}, \boldsymbol{A}, \boldsymbol{\Gamma^u}, \boldsymbol{\Omega^\xi})$, where $\boldsymbol{\Lambda}$ is the matrix of loadings, $\boldsymbol{A}$ is the matrix of coefficient of the VAR in (9), $\boldsymbol{\Gamma^u}$ is the variance-covariance matrix of the error in the VAR and $\boldsymbol{\Omega^\xi}$ is the variance-covariance matrix of the idiosyncratic component in (1).

We run the E-step for k^* times and the M-step for $k^* + 1$ times, until the maximized expected log-likelihood converges and additional iterations stop resulting in significant improvements. The final common factors that will be decomposed into common trend and cycle are those found in the last $(k^* + 1)^{th}$ iteration of the EM algorithm.

2.3 Morley-Wong model specification

2.3.1 Bayesian inference

In this section we analyze the Bayesian VAR (**BVAR**) model. Bayesian inference, differently from the classical frequentist approach where it is assumed that there exist "true" parameter values, treats every parameter of interest as a random variable defined by some probability distribution. In a nutshell, Bayesian analysis consists in combining the prior information the one may have about the distribution for these parameters (**the prior distribution**) with the information contained in the data (**the likelihood function**), in order to obtain a distribution that takes into account the information carried by both these sources of information (i.e. **the posterior distribution)**. This sort of updating is made trough the **Bayes Rule** (Dieppe et al. [2016]):

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)} \tag{10}$$

where θ is the vector of parameters of interest; y is the set of data; $\pi(\theta|y)$ is the posterior distribution of θ given the information in y; $\pi(\theta)$ is the prior distribution; $f(y|\theta)$ is the data likelihood function; $f(y)$ is the density of the data. Bayes rule can be conveniently reduced to:

$$\pi(\theta|y) \propto f(y|\theta)\pi(\theta) \tag{11}$$

Since we are interested in a VAR estimation, the parameters of interest are the vector of coefficients β and the error variance-covariance matrix Σ. So (11) can be rewritten as:

$$\pi(\beta, \Sigma|y) \propto f(y|\beta, \Sigma)\pi(\beta, \Sigma) \tag{12}$$

where $\propto$ stands for "proportional to". In few words, estimating a **Bayesian VAR** consists in specifying a prior distribution for the parameters, together with the hyper-parameter $\boldsymbol{\lambda}$, and then updating the prior beliefs according to the observed data in order to get a posterior distribution over the parameters. Sometimes it is possible to derive the posterior analytically, while for most of the times it is necessary to approximate it through computational methods such as the **MCMC (Markov Chain Monte Carlo)**.

[8] If it is not statistically different from 0, then we do not add a deterministic trend.

2.3.2 Specification of the model and discussion of the Minnesota prior

The model hereby discussed is in the form of Morley and Wong [2020], but some modifications are made in order to make it more suitable for the Euro area analysis Morley et al. [2023]. The whole analysis is built around the **Beveridge ad Nelson decomposition** Beveridge and Nelson [1981], which defines the trend of a time series, y_t, as the long-horizon conditional forecast minus any deterministic drift. The BN trend is defined as:

$$\tau_t^{BN} = \lim_{j \to \infty} \mathbb{E}_t[y_{t+j} - j \cdot \mathbb{E}_t[\Delta y_t]] \tag{13}$$

$\mathbb{E}_t[\Delta y_t] = \mu$ is the deterministic drift or unconditional mean of the growth rate of y_t, while the Beveridge Nelson trend is equal to the conditional expectation only if the unconditional mean of the cyclical component is zero. In this framework y_t is the natural logarithm of the euro area real GDP and the first difference Δy_t approximates the GDP growth rate from one quarter to another, as data are recorded on quarterly basis. Moreover, let $\Delta \tilde{y}_t$ (the demeaned output growth) be the k^{th} element of $\tilde{\boldsymbol{x}}_{\boldsymbol{t}}$, and s_k be a $1 \times Np$ selector vector of zeros with 1 only on k^{th} element. Morley [2002] shows that the output gap of the BN cycle y_t can be computed as:

$$c_t^{BN} = -\boldsymbol{s_k F}[\boldsymbol{I} - \boldsymbol{F}]^{-1}\boldsymbol{X_t} \tag{14}$$

Consider the following VAR(P) model in the companion form:

$$\boldsymbol{X_t} = \boldsymbol{F X_{t-1}} + \boldsymbol{H e_t} \tag{15}$$

where $\boldsymbol{X_t} \equiv [\tilde{\boldsymbol{x}}'_{\boldsymbol{t}}, \tilde{\boldsymbol{x}}'_{\boldsymbol{t-1}}, ..., \tilde{\boldsymbol{x}}'_{\boldsymbol{t-p}}]'$, $\tilde{\boldsymbol{x}}_{\boldsymbol{t}}$ is a $N \times 1$ vector of demeaned variables with $\tilde{\boldsymbol{x}}_{\boldsymbol{t}} \equiv \boldsymbol{x_t} - \boldsymbol{\mu}$, $\boldsymbol{\mu}$ is $N \times 1$ vector of means of $\boldsymbol{x_t}$, $\boldsymbol{F}$ is a companion matrix , $\boldsymbol{e_t}$ is a $N \times 1$ vector of forecast errors and finally $\boldsymbol{H}$ is a matrix mapping the forecast errors to the companion form. The idea is to estimate a BVAR casting it into the form implied by (9), and then to use the Beveridge Nelson (**BN**) decomposition in order to construct the output gap. We specify the BVAR for demeaned variables as:

$$\tilde{\boldsymbol{X}}_{\boldsymbol{t}} = \boldsymbol{\Phi_1}\tilde{\boldsymbol{x}}_{\boldsymbol{t-1}} + ... + \boldsymbol{\Phi_p}\tilde{\boldsymbol{x}}_{\boldsymbol{t-p}} + \boldsymbol{e_t} \tag{16}$$

In matrix form:

$$\tilde{\boldsymbol{X}}_{\boldsymbol{t}} = \begin{bmatrix} \phi_1^{11} & \dots & \phi_1^{1n} & \phi_2^{11} & \dots & \phi_2^{1n} \dots & \dots & \phi_p^{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \phi_1^{n1} & \dots & \phi_1^{nn} & \phi_2^{n1} & \dots & \phi_2^{nn} \dots & \dots & \phi_p^{nn} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{X}}_{\boldsymbol{t-1}} \\ \tilde{\boldsymbol{X}}_{\boldsymbol{t-2}} \\ \vdots \\ \tilde{\boldsymbol{X}}_{\boldsymbol{t-p}} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ \vdots \\ e_{n,t} \end{bmatrix} \tag{17}$$

$\boldsymbol{E}[\boldsymbol{e'_t e_t}] = \Sigma$ is the variance-covariance matrix of the errors and $E[\boldsymbol{e'_t e_{t-i}}] = 0 \; \forall i > 0$, meaning that errors are uncorrelated for every previous period. Also, the variables are demeaned using their sample average.

In the Bayesian framework the prior distribution contains the prior beliefs of the econometrician about the parameters of interest. Once the beliefs on the parameters are expressed, one wants to determine how much the prior beliefs are sensible to the observed data and this is made by the means of the hyperparameter λ. As $\lambda \longrightarrow 0$, the posterior parameters will stay closer to the prior beliefs, meaning that the variables of the VAR are independent white noise processes. On the contrary, as λ increases the data evidence will prevail over prior beliefs. A more soft prior will likely come with an higher value of λ and viceversa. A right choice of λ is crucial for the performance of the model, and even small differences will affect significantly the results.

The **Minnesota-type prior** applies the shrinkage to the VAR parameters as follows:

$$\mathbb{E}[\phi_i^{jk}] = 0 \tag{18}$$

$$Var[\phi_i^{jk}] = \begin{cases} \frac{\lambda^2}{i^2}, j = k \\ \frac{\lambda^2}{i^2}\frac{\sigma_j^2}{\sigma_k^2}, otherwise \end{cases} \tag{19}$$

where ϕ_i^{jk} is the slope coefficient of the i^{th} lag of variable k in the j^{th} equation of the VAR in the matrix form of equation (17). The variances of σ_j^2 and σ_k^2 are unknown and estimated by the variance of the residuals of an **AR(4)**.

Because of the inclusion of the foreign sector in the model, further deepened in the next paragraph, our estimation is based on a variation of the Minnesota prior, namely the **Normal independent Wishart prior**, just like Morley et al. [2023]. The benefits of this prior comes with the price that the joint posterior distribution cannot be analytically derived and it must be approximated through Monte Carlo methods. Indeed, as the conditional posterior distributions can be analytically derived, the joint distribution can be approximated by the means of MCMC methods such as the **Gibbs sampler**.

In our estimates, we choose a value for λ of **0.2**, which has been shown to be a reliable value in the BVAR estimation literature Carriero et al. [2015]. Although Morley et al. [2023] affirm that this value results suitable in analysing pre pandemic samples, we also estimate the model with different λ in section (4.3) of the empirical results.

2.3.3 Model modification due to foreign sector

In our model, a block exogenous structure on the coefficients is added, accounting for a foreign sector. Hence, the model has a two-block structure with a foreign and a domestic blocks. We impose this kind of structure when we believe that some variables determining the model are exogenous to it, in the sense that they are not influenced by the other variables of the model itself. In this specific context one can state that the Euro area is believed to be at least partially determined by foreign economies, but, on the other hand, is not able to have an effect on the latters. Including such a structure allows us to capture the effect of the foreign sector on the Euro area, allowing for more accurate analysis and predictions.

This modification is not harmless in terms of estimations, as posterior moments of the model cannot be calculated analytically anymore and we thus need to rely on approximations methods for the posterior distribution. This induces a huge cost in terms of computational efficiency, but given the better estimation results this trade-off is accepted. The VAR can be rewritten as:

$$\begin{bmatrix} z_t^* \\ z_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi_1^{11}} & \mathbf{0} \\ \boldsymbol{\Phi_1^{21}} & \boldsymbol{\Phi_1^{22}} \end{bmatrix} \begin{bmatrix} z_{t-1}^* \\ z_{t-1} \end{bmatrix} + ... + \begin{bmatrix} \boldsymbol{\Phi_p^{11}} & \mathbf{0} \\ \boldsymbol{\Phi_p^{21}} & \boldsymbol{\Phi_p^{22}} \end{bmatrix} \begin{bmatrix} z_{t-p}^* \\ z_{t-p} \end{bmatrix} + \begin{bmatrix} \boldsymbol{A_{11}} & \mathbf{0} \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} \end{bmatrix} \begin{bmatrix} \epsilon_t^* \\ \epsilon_t \end{bmatrix}$$

The vector $\boldsymbol{\tilde{x}}$ is partitioned into $\boldsymbol{\tilde{x}_t} = [\boldsymbol{z^{*\prime}_t}, \boldsymbol{z_t^{\prime}}]'$, $\boldsymbol{z_t^*}$ and $\boldsymbol{z_t}$ are both vectors of demeaned variables respectively of dimension $N^* \times 1$ and $[N - N^*] \times 1$ of foreign and domestic variables.

2.4 Morley-Wong estimation

The restrictions that we made through the block-exogenous structure forbid us to exploit the natural conjugate prior as used by Morley and Wong [2020], forcing us to rely on estimation methods for approximation of the posterior distribution like Morley et al. [2023]. We thus used a Markov chain Monte Carlo (MCMC), which is used to approximate the posterior joint distribution. This particular MCMC method consists in a Gibbs Sampler, where the prior distribution is Minnesota type with an independent normal-Wishart structure. The Gibbs sampler is used when one wants to draw from a multivariate distribution, but it is able to draw only from the conditional distribution and not from the joint one.

The parameters of the posterior distribution are initialized by **OLS** values and then the Gibbs sampler is performed. In our analysis we extract 15000 draws; from this 15000, the first 7000 are "burned" and only the remaining 8000 are kept in consideration. This because, since the samples drawn from the conditional distribution constitutes a Markov Chain, it may take some time for the Markov Chain to converge to its stationary distribution, which can be shown to be our joint distribution of interest. Hence, we discard a certain number of draws to be sure of drawing from the joint distribution despite that from the conditional. Once we have obtained the draws, those can be used to build an empirical posterior distribution.

2.5 Hodrick–Prescott (HP) filter

The HP filter is a standard and widely used method for the estimation of the output gap. In practice, it allows to extract the trend and the cycle from a generic time series, hence permitting to estimate the potential output and the output gap. This method requires to choose an hyperparameter $\boldsymbol{\lambda}$. In our case, we pick a value of **1600** for λ, which is very widely used in the literature for quarterly data. As a matter of fact, it performs very similarly to the Morley et al. [2023] model. By adjusting the value of λ, the degree of smoothness applied to the time series can be controlled. The higher is λ, the smoother is the trend with less fluctuations, while a lower value of λ consents the trend to follow more closely the original data.

3 Data description

3.1 Data set

For both models we used the same dataset, composed of quarterly European data for 16 time series from Q1 1999 to Q4 2021, for a total of 92 observations for each time series. The 16 time series are divided into 2 groups, a domestic and a foreign block. The domestic block is composed by: euro area real GDP, industrial production, employment, housing permits, CPI, policy rate, hours worked, term spread, capacity utilization, unemployment rate, PMI, risk spread, real effective exchange rate and household debt over GDP. The foreign block instead is composed by US real GDP and the real price of oil. In table 1 we provide detailed information regarding the data sources.

15 out of the 16 time series used are the same that were composing Morley and Wong [2020] dataset. On top of them, we add a series for the household debt over GDP ratio, in order to further disentangle financial data. Household debt often serves as a critical economic indicator because it reflects the financial health of individuals and can influence broader economic trends. Accounting for the household debt in the analysis can provide insights into monetary policy effectiveness, financial stability and overall economic performance. For instance, a negative relation between household debt and output growth have been shown (Alter et al. [2018]). When used, this variable is multiplied by GDP in order to eliminate the ratio and to express it in Millions of euros.

The dataset is also composed by institutional estimates of the output gap for the Euro Area, made by the OECD, the European Commission, the IMF (World Economic Outlook database) and the ECB (using an Unobserved Components Model). While the UCM estimate is based on quarterly data like our estimations, OECD, IMF and EC estimations are based on annual data, and so they might not be fully comparable to our quarterly estimates.

For further analysis over the Barigozzi and Luciani [2023] model, we also add other 12 financial variables into our dataset, all listed in table 2. Financial information seem to be important to obtain meaningful estimates of the business-cycle, as noted by Borio et al. [2017], even though the larger driver of this phenomenon has been shown to be households' debt, already included in our original dataset.

3.1.1 Foreign Block

Following Morley et al. [2023], we use a standard two-block foreign and domestic structure, thus considering in our dataset also non European data, while fitting their model. This is made to address the fact that the euro area is a relatively open economy and thus US economic factors can have an influence over European variables.

Description	Units	Source	T. BL	T. MW
U.S. real GDP	Billions of chained 2012 Dollars	Federal Reserve Economic Data (FRED)	1	2
Real brent oil price	US-dollar $\times$ Ratio (Euro/US-dollar)	World Bank	1	1
Euro area real GDP	Millions of chained 2015 EUR	Eurostat	1	2
Industrial production index (excl. construction)	Index 2010 = 100	OECD	0	2
Employment	Thousands of persons	AWM database and Eurostat	1	2
Harmonised index of consumer prices (HICP)	Index 2010 = 100	Eurostat	3	2
Building permits: residential buildings	Index 2010 = 100	Statistical Office of the European Communities	0	2
Nominal short-term interest rate	Percent	Eurostat	0	3
Hours worked	Index 2010 = 100	Eurostat	0	1
Euro area 10-year government benchmark bond yield minus Euro 3-month Libor	Percent	ECB and Thomson Reuters	0	0
Capacity utilization in manufacturing	Percent	Eurostat	0	0
Unemployment rate	Percent	Eurostat	0	0
PMI Manufacturing output	50 = no change on previous month	Markit	0	0
Euro area credit risk spread	Index	Gilchrist and Mojon [2018]	0	0
Real effective exchange rate	Index	FRED	1	1
Household debt	Eurostat	Million EUR	1	2

"T. BL" = Transformations used for the Barigozzi-Luciani model.

"T. MW" = Transformations used for the Morley-Wong model.

Transformations: 0 - Level; 1 - (Log level)*100; 2 - (Log differenced)*100; 3 - Differenced.

Table 1: Data sources

Description	Units	Source	T. BL
Total Economy - Total financial assets	Million euro	Eurostat	1
Total Economy - Total financial liability	Million euro	Eurostat	1
Total Economy - Debt securities	Million euro	Eurostat	1
Total Economy Loans	Million euro	Eurostat	1
Non Financial Corporations - Total financial assets	Million euro	Eurostat	1
Non Financial Corporations - Total financial liabilities	Million euro	Eurostat	1
Non Financial Corporations - Loans	Million euro	Eurostat	1
Household - Total financial assets	Million euro	Eurostat	1
Household - Loans	Million euro	Eurostat	1
General Government - Total financial assets	Million euro	Eurostat	1
General Government - Total financial liabilities	Million euro	Eurostat	1
General Government - Debt securities	Million euro	Eurostat	1

Table 2: Financial data sources

We thus treat the foreign sector as being block-exogenous with respect to the Euro area, similar to the traditional small open economy literature (Zha [1999]; Justiniano and Preston [2010]; Kamber and Wong [2020]), hence imposing the top-right block $N^* \times N$ block of the ϕ_j matrix to be zero. The block-exogeneity identification restriction imposes that the domestic economy is too small to affect the foreign economy, namely, euro area variables do not enter as explanatory variables the equations within the foreign-country block, both contemporaneously and through lags. This is quite a realistic assumption, given the structure and the dimension of the US economy, which can be considered approximately closed; hence, the influence of the domestic variables over the US economy can be assumed to be negligible.

While addressing the Barigozzi and Luciani [2023] model, instead, we do not impose such a restriction over our data. Foreign variables are still considered since they can have quite a big influence over euro area economy. However, notice that we also estimate this model without considering the foreign variables, to see whether the neglected block-exogeneity assumption would bring to wrong estimates. Indeed, as shown by Zha [1999], not taking adequately into account the small economy assumption can lead to anomalous results. We thus discuss this point in the results section while comparing the two outgap measures resulting from the two models.

3.1.2 Covid Data

The time series we collected contain also data relative to the Covid-19 pandemic time period. While the BL[9] model do not contain a specific approach to address those data, also because the dataset analyzed in the original paper collects data up to 2019, the MW[10] model instead directly analyzes pandemic data following the Lenza and Primiceri [2022] specification, in which the residual covariance matrix in the BVAR model is scaled up by a factor (a factor that has a different values for 3 different parts of the pandemic period and takes value 1 outside the Covid-19, so to make the specification to collapse back into the standard BVAR model).

Pandemic period observations can be considered mainly as outliers with respect to the other data and those outliers can induce nontrivial changes to the estimated parameters from a BVAR. Thus, while using the MW model,

[9] Barigozzi and Luciani [2023] model
[10] Morley et al. [2023] model

we should have followed this specification as well if we were using also pandemic data. However, in order to make the two models as much comparable as possible, we decide to exclude Covid-19 data from our sample, thus avoiding using the Lenza and Primiceri [2022] specification when running the MW model (also making it computationally less heavy). We thus retain observations up to Q4 2019 included[11], ending up with 84 observations for each time series.

3.2 Preliminary transformation

Preliminary transformations are used to induce raw data to be stationary, when needed, consistently with the existing literature (Stock and Watson [2012]; McCracken and Ng [2016]). For both models, series are subject to a preliminary screen for outliers[12] and then transformed to reduce their integration order. The particular preliminary transformations are model specific, hence we discuss them separately. Pertaining our replication of Morley and Wong, while keeping interest rate spreads in level, we take the first difference of logs for real activity variables, thus transforming them into quarterly growth rates, and take first differences for the policy rate. Prices, hours worked and exchange rates are taken in logs. In the tables above regarding the data sources all the transformations are listed. Concerning our emulation of the BL model, we apply a first difference transformation for selected variables such as price indicators, similarly to Stock and Watson [2016], and take logs of other variables appropriately listed in the tables(1) and (2) above. This different approach is pursued because the BL model can handle both I(0) and I(1) dynamics and thus we need to transform only those variables which exhibit I(2) dynamics. Mirroring the BL approach, variables are scrutinized through a test that identifies which ones must be detrended; if the test yields a positive result for a given series then detrending occurs for said series. If variables are not detrended, implying they are not flagged by the forenamed test, then they are demeaned instead.

4 Empirical Results

4.1 Model comparison and role of foreign and financial variables

All the estimations performed using the Barigozzi-Luciani approach are in Figure 1. We use different set of series in order to test the model sensitiveness to different variables. One main difference in estimating the output gap for the Euro Area rather than for the US (like in Barigozzi and Luciani [2023]) regards the assumption of not accounting for the role of foreign economies. In our estimations, hence, we will perform the analysis both when adopting this assumption and when allowing for the effect of EA main trading partner (the US), as well as for its integration in the global economy. Starting from the top of the figure, panel (1a) and (1b) display the output when taking into account only domestic macro-variables with the addition of household's debt as the only financial variable. In panel (1c) and (1d) we add the two foreign variables (i.e. US GDP and Oil price), on top of the aforementioned ones. Moving on, to estimate models displayed in panel (1e) and (1f), we considered only domestic variables (excluding US GDP and Oil price) while adding financial variables listed in Table 2 (on top of household debt). Finally, panel (1g) and (1h) report the output gap estimated using all the series we collected (i.e. domestic, foreign, macro and financial). Taking the UCM and Wong measures as benchmark, we can see that when not accounting for the financial variables, adding the foreign ones results in a marginal improvement of our estimate. We notice, however, that this holds only up to the 2011 (see panel (1c)). The reason could lie in the fact that the model we are using is built for the US and not accounting for foreign economies (i.e. a large and, in the model, closed economy). If we add them without changing the structure of the model, which would enable it to take into account that they are specifically foreign variables (and in this respect are different to all the others), then we would be using the wrong variables for a given model. Indeed, once we account for financial variables (panels (1f) and (1h)), adding the

[11] We consider the first confirmed Covid-19 infection in Europe (January 2020) as the starting point of the pandemic period. We thus retain all the quarterly data strictly before that date.

[12] Following McCracken (2021) we use here the definition of outlier as an observation that deviates from the sample median by more than ten interquartile ranges.

foreign ones even worsens our estimates. On the other hand, adding financial variables *per se* results in a significant improvement of our estimate of output gap (compare panel (1a) to panel (1f)). Our hypothesis is that, hence, adding foreign variables while not accounting for financial ones we are implicitly adding information which partially capture financial dynamics. For this reason the estimate marginally improves. Adding them to the model when we do account for financial variables (which proves to be important), instead, worsens the performance. In this sense, we could say that foreign variables only introduce noise to the model. Overall, hence, the estimate displayed in panel (1f) is the one closer to our benchmark. In particular, it captures the correct dynamics, while it fails to get correct numerical estimates. The reason for this stems from the fact that consistency of the estimated factors needs a large number of series (n very large). In our case, hence, this is clearly not achieved being n, at most, equal to 28.

We conclude this subsection on a technical note, comparing the efficiency of the two codes in terms of time taken to perform the calculations using the MATLAB programming language. The BL model is significantly faster to estimate than the Morley and Wong one. This is the case because the former only takes a few minutes to run and yield its estimates, and this remains pretty consistent whilst increasing the number of series. On the other hand, Morley and Wong estimates can be computed pretty quickly with a small number of draws, but require a huge amount of these in order to pinpoint quantitatively accurate estimates. And together with precision, also computational intensity and thus time taken to run the code increases dramatically with the number of draws, taking many hours to properly run.[13] Therefore, in instances were time or computing power is a constraint, the computational efficiency of the BL code, which reflects an efficiency in signal extraction from the data, might be a relevant factor when evaluating which model to use for the estimates.

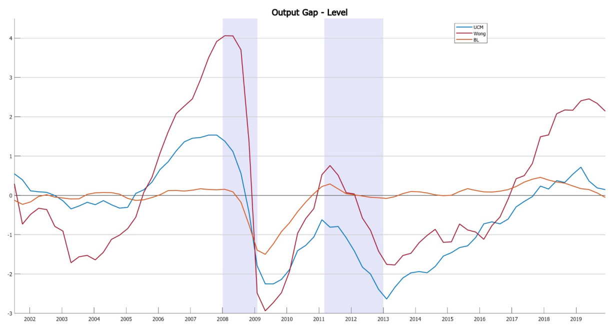
4.2 Morley and Wong baseline estimates

The top panel in figure 2 presents the estimated output gap for our 16 variables, using $\lambda = 0.2$. Overall, the output gap estimated, which is quite similar to the one estimated by Morley et al. [2023], is aligned with all the turning points of the business cycle as dated by CEPR ever since the monetary union came into existence. It is also in line in representing the two main crisis of the last two decades, the "Great Recession" of 2008-2009 and the European sovereign crisis in 2010-2012, shown in the graph by gray bands.

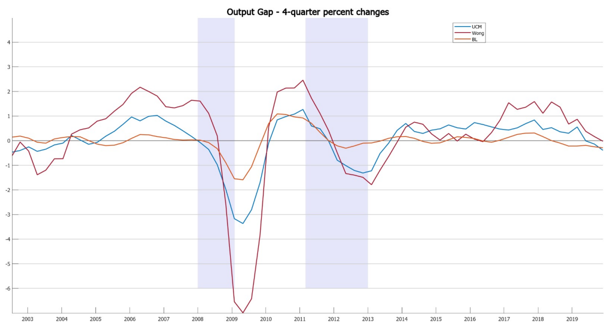
The bottom panels instead report the comparison between our estimates and all the other estimates we used as benchmarks. Institutional estimates by the OECD, European Commission and the IMF (bottom-left panel) assume a production function for the total economy and define the output gap as the difference between actual output and the estimated trend according to the production function. Notice that these measures are based on annual data, so they are not completely comparable with our estimates. The other estimates (bottom-right panel) are instead based on quarterly data: UCM refers to estimates for an unobserved components model by ECB, the HP filter estimate is constructed with its typical smoothing parameter of 1600. The HP filter estimate results to be the more consistent with our results, while institutional estimates give an output gap that is quite different for all the year of the series, with estimates consistently lower than ours after the 2008 crisis.

Given that the HP filter estimates are quite similar to ours, we can than say that the main advantage of the Morley et al. [2023] model is given by the chance of explaining the informational content of single variables, as discussed further in section 4.4.

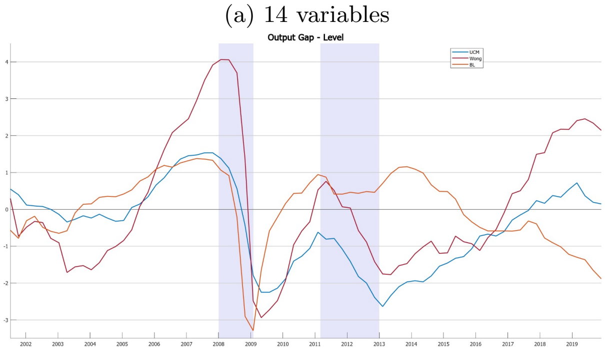
[13]Indeed, in our final run of the code with 15 thousand draws and 7 thousands burn-in it took us 8 hours to get the estimates. Time taken to make computations could be significantly reduced if one were interested only in the estimation of the baseline model, which would, however, still take a significant amount of time to run.



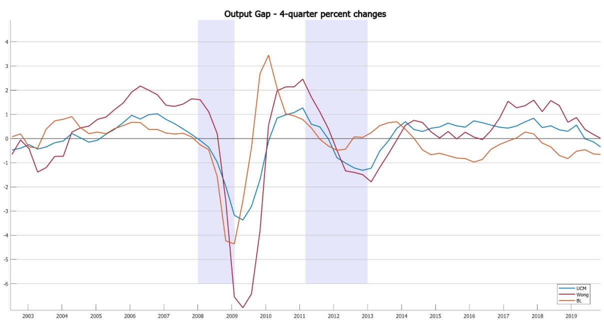
(a) 14 variables



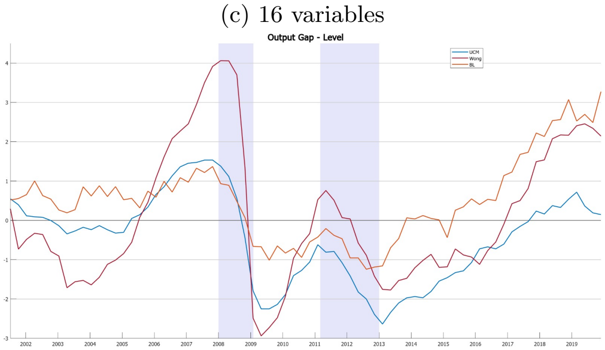
(b) 14 variables: Year-over-year growth



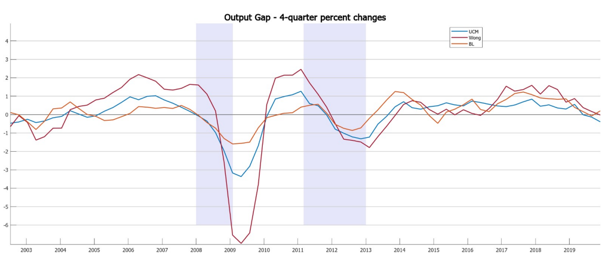
(c) 16 variables



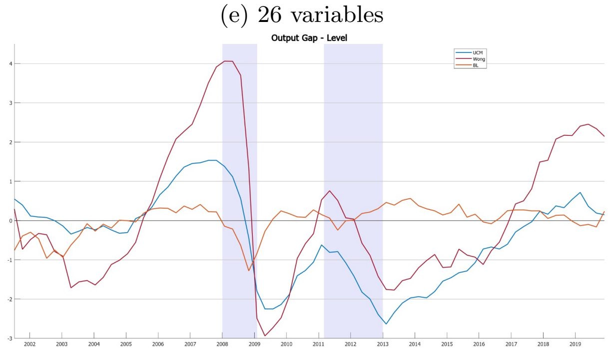
(d) 16 variables Year-over-year growth



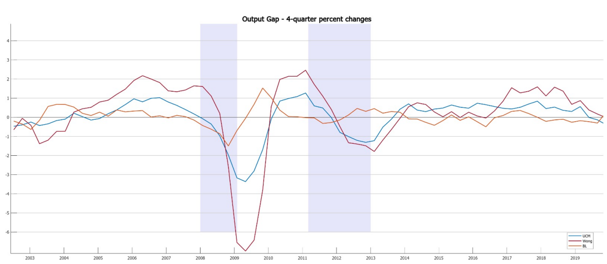
(e) 26 variables



(f) 26 variables Year-over-year growth



(g) 28 variables



(h) 28 variables Year-over-year growth

Figure 1: Comparison BL, MW and UCM

4.3 Effects of the shrinkage hyperparameter

In figure 3 we explore the effects of changing the value of the shrinkage hyperparameter over the output gap estimates. Following Morley et al. [2023], three different values of λ are considered: 0.075, 0.2, 0.75, 0.9, where 0.2 is the value we used for our baseline model fit, 0.75 is the value the authors found to be the optimal when considering Covid-19 data and the other two are just comparison measures. More shrinkage, so lower λ, reduces the amplitude of the output gap, at least for the first half of the sample period, because the higher the shrinkage

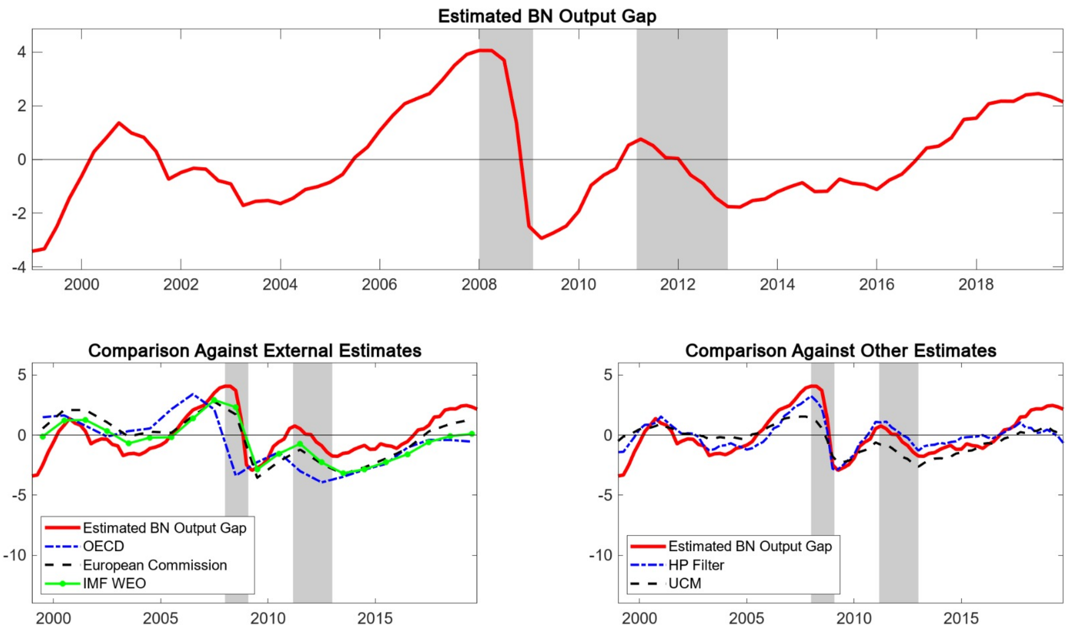


Figure 2: Wong baseline estimates

(smaller λ) the less important is the cyclical component. Relaxing the shrinkage, instead, amplifies the gap. Overall, no big changes in the estimates are noticeable for different λ values; however, we can see that if we would have used $\lambda = 0.75$ as in Morley et al. [2023] without including Covid data, the output gap estimate would have been slightly overestimated for the period before the financial crisis and underestimated for the period after the sovereign debt crisis.

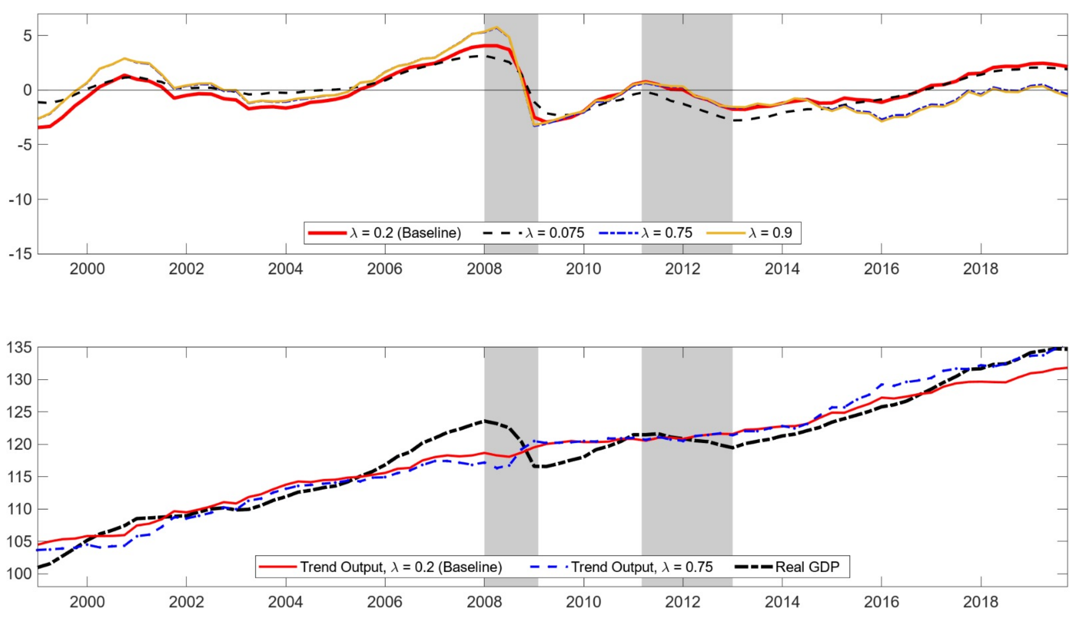


Figure 3: Estimates for different shrinkage hyperparameters

4.4 Informational content of variables

As briefly discussed in the first section of the empirical results, one of the benefit of the multivariate analysis is that we can decompose the output gap into sources of forecast errors, separating all the variable used. Figure 4 shows the standard deviations of the informational contributions for each of the 16 variables in the baseline model, providing useful clues as to which variables have been more important on average. Contributions are calculated based on an informational decomposition of the estimated output gap into different types of forecast errors. IP is industrial production, CAPU is capacity utilization, PMI is the purchasing manager index for manufacturing output, and RER is the real exchange rate.

Except from euro area GDP, IP, Policy rate and PMI, almost all the variables seem to contribute in a non negligible way to the estimation of the output gap. We see that, as in Morley et al. [2023], Hours worked results to be one of the most important variables in explaining the output gap, in contrast to the findings of Morley and Wong [2020], where for the estimation of the US output gap they found that Unemployment was more important. We also notice that the foreign block variables account for a large share of the information content, mostly the real oil price. This data confirm the importance of considering the influence of foreign variables and highlighting the relevance of considering an open-economy setting in the analysis of European data.
In contrast to Morley et al. [2023], few variables, such as employment and housing permits, seems to be less informative in our sample estimates. Two are the possible explanations of this finding: one possibility is that those variable are relatively more important in the estimation for the output gap during the Covid-19 period, and, since we are not including Covid data in our analysis, the informational content is reduced. Similar conclusion have been made by the authors in the original paper. The second possible explanation is that adding the Household debt variable, part of the informational content of the other variables is replaced inside Household debt information. This possibility seems quite reasonable, given that Household debt accounts for a significant part of the information content. In light of this, we can confirm our hypothesis for which adding the Household debt variable can be quite useful in order to account for the role of financial variables in the model estimates. Hence, the role of financial variables appears to be not negligible in the setting.

Pertaining (5), we can note a few things. To begin, in the top panel our baseline estimates are compared to an AR(4). The latter doesn't provide much informative content and fluctuates around 0 for most of the time spanned by our series. One notable exception occurs during the 2008 crisis, where the AR(4) captures with a year of lag, at the end of the recession, the highly positive output gap that precedes the financial crisis. In contrast, the baseline incorporates this information in a timely fashion. A key fact to notice is that in Morley et al. [2023] the AR doesn't show this positive spike; we attribute this difference to our choice of including the household debt series. Indeed, it seems reasonable to assume that the aforementioned spike is due to a lagged effect of the peak in household debt, a key factor in the 2008 financial crisis. Moreover, in the bottom panel of (5) the baseline estimates are compared with fits of the model performed using different subsets of variables. The 12 variable subset is obtained by dropping IP, RER, PMI and unemployment, whilst the other subsets names are self-explanatory 'drop real exchange rate', 'drop term spread'. The baseline estimate is largely similar to the one performed using the subset comprising of 12 variables and the one obtained by dropping term spread. The 12 variable estimate seems to overshoot the positive output gap observed prior to the 2008 crisis and yields a mildly smaller positive output gap towards the end of sample. On the other hand, dropping term spread leads to a higher estimate of the output gap towards the end of sample with respect to the baseline but yields similar results otherwise. Although qualitatively similar, these three estimates differ in their exact quantitative results, this is a notable difference from Morley et al. [2023] paper, in which these three estimates and the one obtained by dropping the real exchange rate, which we shall separately discuss, basically overlap. We attribute this discrepancy to our restricted time sample; this is the case since the issue persists also when dropping household debt and performing similar estimates. The same explanation applies when we try to analyze the discrepancy, which is not present in the reference paper, between the baseline and the 'drop real exchange rate' estimates. This is the case since the mismatch persist when estimates are made, ceteris paribus, without the household debt series. Indeed, with our data vintage, dropping real exchange rate yields estimates that are starkly different from the relatively homogeneous ones yielded by the other estimates, both qualitatively and quantitatively. We avoid discussing the discrepancy in detail since the estimates obtained by dropping the real exchange rate seem to simply not provide any meaningful information neither in estimating nor analyzing the variable of interest. In other words, they seem to be too influenced by noise and follow no specific economic logic. The fact that by truncating the data at the end of 2019 this discrepancy between estimates arises is a puzzle of its own. We do not have a clear cut answer to why this happens, and the phenomenon might as well be just a statistical curiosity or glitch. However, were it not the case, further analysis might be needed to shed light on the topic.

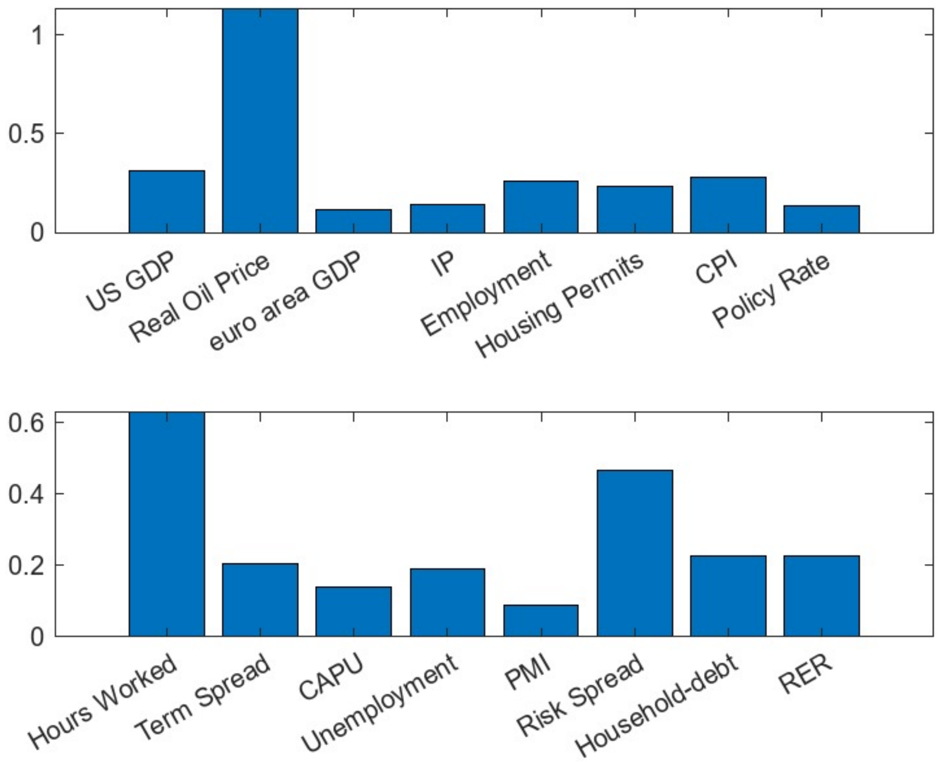


Figure 4: Variables' informational contents

4.5 The role of hours worked

In figure 6, we look at an historical informational decomposition of the output gap, presenting the share of the forecast errors of hours worked separated from the other 15 variable grouped together. The top panel reports the estimates for a value $\lambda = 0.2$, so the value used in our baseline model, while the bottom panel is characterized by a value of $\lambda = 0.75$, as the one that Morley et al. [2023] found to be the optimal while analyzing Covid data. First notice that the estimates of the two panel differs notably, reporting opposite information about the contribution of hours worked to the output gap for almost all the time period. Given that Morley and Wong in their paper found quite similar results comparing both values of λ, we can presume that in our setting, so the one where Covid data are not included inside the dataset, the results obtained fitting the model with $\lambda = 0.75$ are misleading. A bigger value of the shrinkage hyper parameter can be the optimal choice while addressing the Covid-19 pandemic, but not in our case, and thus we can concentrate on discussing the top panel results.

Throughout all the sample period hours worked plays a big role in determining the output gap, except for the central period in which the information content is much lower. Up to the 2008 crisis, hours worked contribute in the same direction of the overall output gap, while after that the trend is reverted.

In figure (7) the time series plot for hours worked, industrial production growth, unemployment rate and capacity utilization are shown. This is done largely to conform with Morley et al. [2023]; however, in that circumstance the authors stressed the pivotal role of hours worked in explaining the output gap during the covid crisis. Since

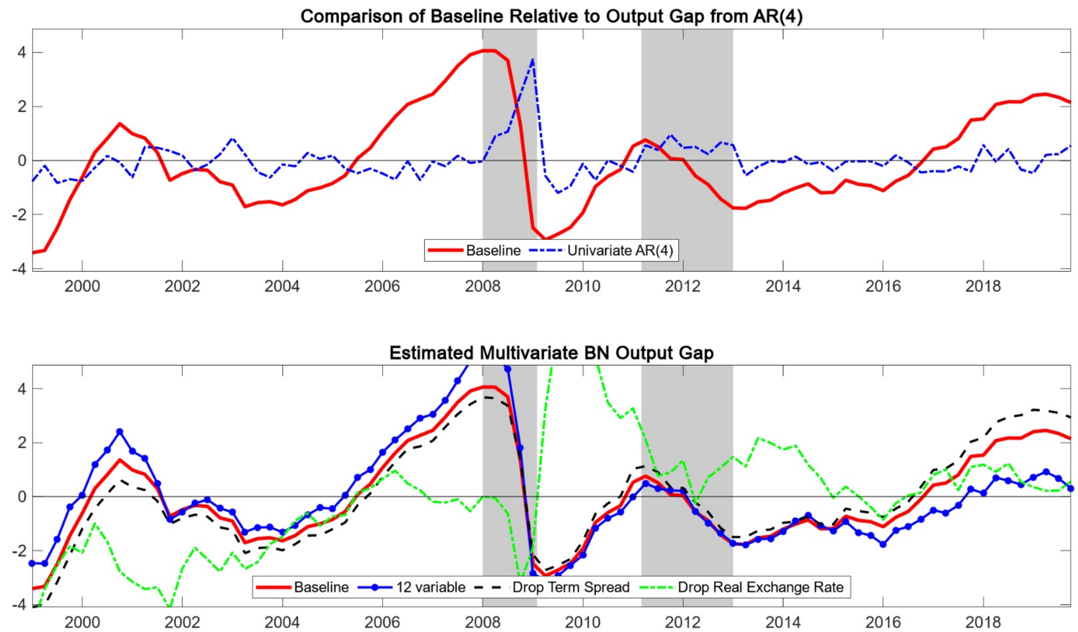


Figure 5: Estimated euro area output gap for various-sized models

we don't consider said data, this figure is of lesser importance to us. Nonetheless, a few interesting remarks can be made. Whilst hours worked doesn't play as a pivotal role in estimating output gap, it is still very relevant and its behaviour is largely coherent with the baseline estimates of output gap. In fact, it peaks before the 2008 crisis then it goes down, recovers and is rising towards the end of sample, like the baseline. On the other hand unemployment rate doesn't seem to behave in a similar fashion nor carry the same informative content, contrary to what is empirically shown to be the case in the U.S economy. Likely, the difference in the degree of labor market frictions between the two economies determines this. Indeed, the unemployment rate seems to be adjusting in a sluggish manner, with the effects of the 2008 crisis manifesting themselves slowly up until 2012, exacerbated by the contemporaneous crisis of the European sovereign debt. In fact, this is coherent with the existent literature on the topic[14], which also shows how firms respond to the crisis by adjusting along the intensive margin rather than the extensive margin. In other words, they prefer, or are forced to reducing the amount of hours worked rather than laying off workers altogether, this further confirms the importance of hours worked in the euro area, coherently with the Figure(4) informational content decomposition.

Pertaining industrial production growth and capacity utilization, we have a quintessential manifestation of the effects of a financial crisis. Indeed, slack in the economy is accrued due to the limited ability of the financial sector to properly support the rest of the economic system needs, leading to under utilization of the present productive capacity and to the plummeting of the industrial production growth which is negative for a similar reason. In other words, production capacity is present but not used due to the financial instability and the presence of uncertainty, which contrary to risk, cannot be insured against and creates disruption of its own kind. In particular, a channel through which this disruption takes place is the reduction in investment commonly associated with financial crisis; which is reflected in business sentiment indicators such as the PMI which plummeted during the financial crisis and recovered only in 2010.[15]

[14] Please see Burda and Hunt [2011] and Ohanian and Raffo [2012] for further information on the topic

[15] The reduction between June 2008 and March 2009, the lowest point is of 40%

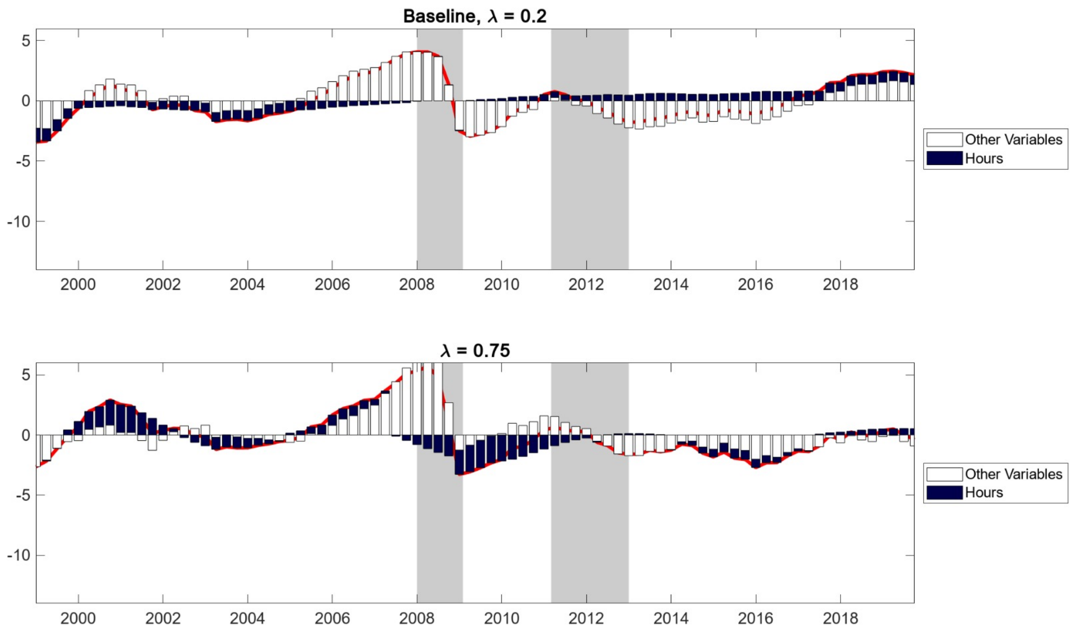


Figure 6: The role of hours worked

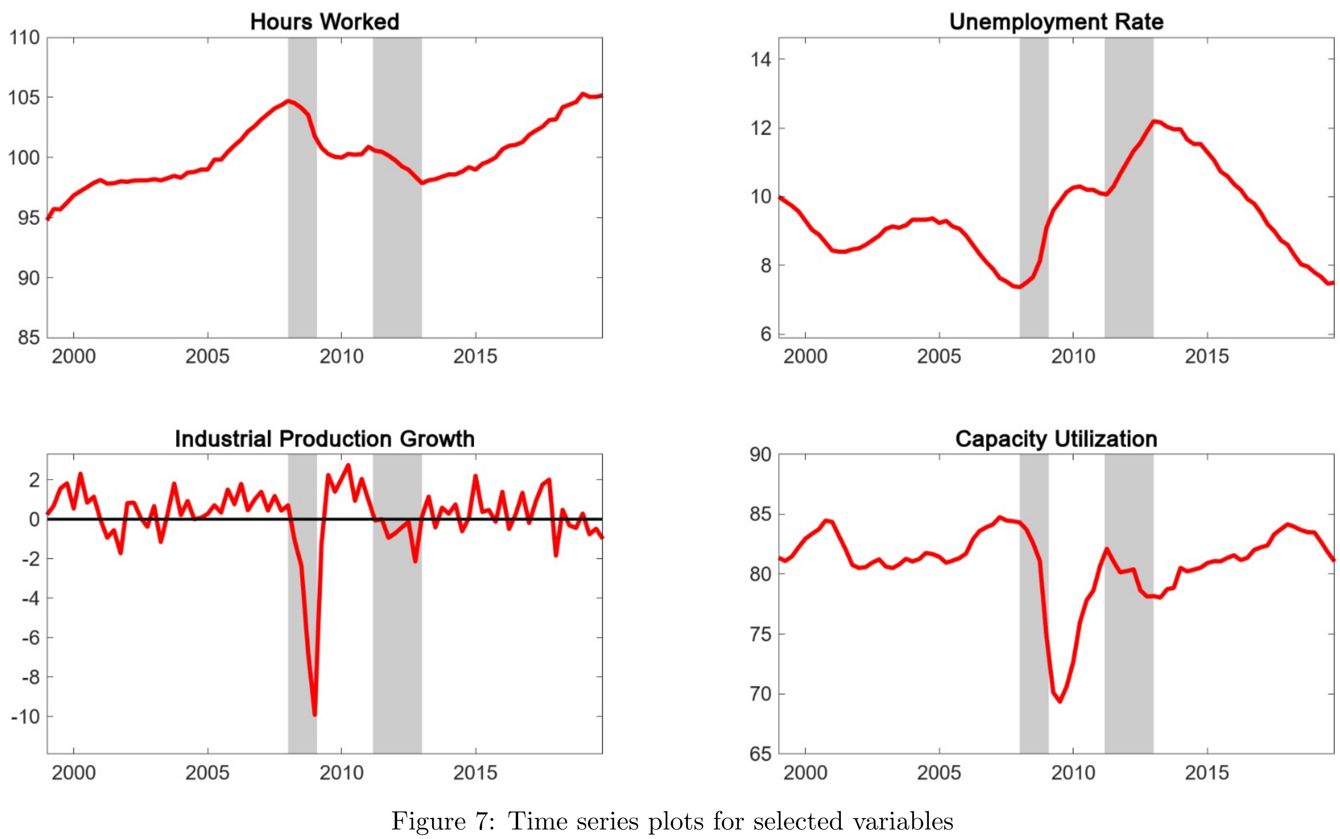


Figure 7: Time series plots for selected variables

5 Conclusions

We measured the Euro area output gap using two different methods, relying on a relatively limited dataset. The BVAR model appears to give results closer to institutional estimates and the HP filter estimate, thus suggesting to be more reliable in a setting with few time series. The DFM, although being computationally more efficient, gives less precise estimates. However, we find that this lack of precision could be assessed by increasing the number of time series used.

As predicted, financial variables appear to have a non negligible influence in predicting the Euro area output gap. We also confirm that household debt contains a big part of all the financial variables information.

Our findings confirm that hours worked has a prominent role in the estimate of the output gap, and a superior informative content compared to the unemployment rate. This is in contrast to what happens in the U.S economy, where the converse is true.

A Appendix

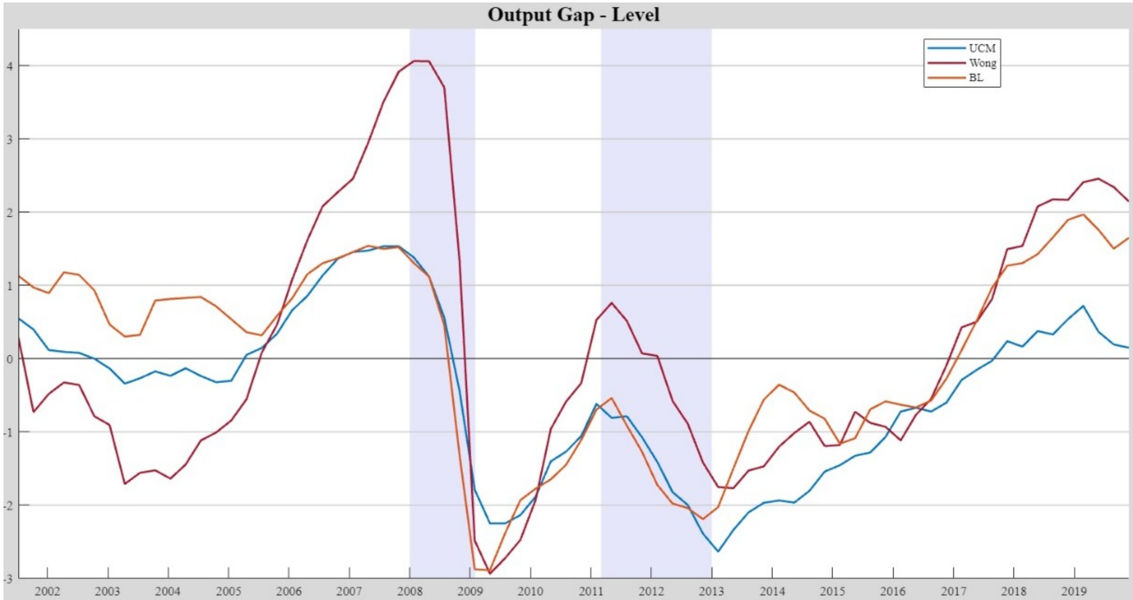
A.1 BL model factors discussion

Throughout the paper, we followed Barigozzi and Luciani [2023] in determining the number of factors q to be used for the model's estimations, hence using $q = 4$ when comparing the results with Morley et al. [2023]. However, we also run some preliminary checks with $q = 2, 3, 4$, verifying that our choice was actually the best in that basket of q_S. Given that this was not the focus of the study, we employed an heuristic approach to evaluate the optimal number of factors. For each q we evaluated the fit of the model by checking its proximity to the two benchmarks. We stopped at $q = 4$ largely to conform with the reference literature, since deviating from it would require employing a formal procedure which was beyond the scope of this paper. Nonetheless, one might investigate whether or not an additional factor might help in improving the model at hand. Evidence for the need of an additional factor comes from the worsened performance of the BL model when US variables are employed in the estimation. However, such poor performance is not due to irrelevance of the additional data. The US economy is both a global economic powerhouse and one of the leading EU trading partners. Therefore, there is need to properly take into account the influence of foreign economies on the European one through an ad hoc factor. This procedure allows to leverage more information in the process of pinpointing the output gap, which is crucial in our setting due to the limited availability of the data, both in terms of the number of series and their length.

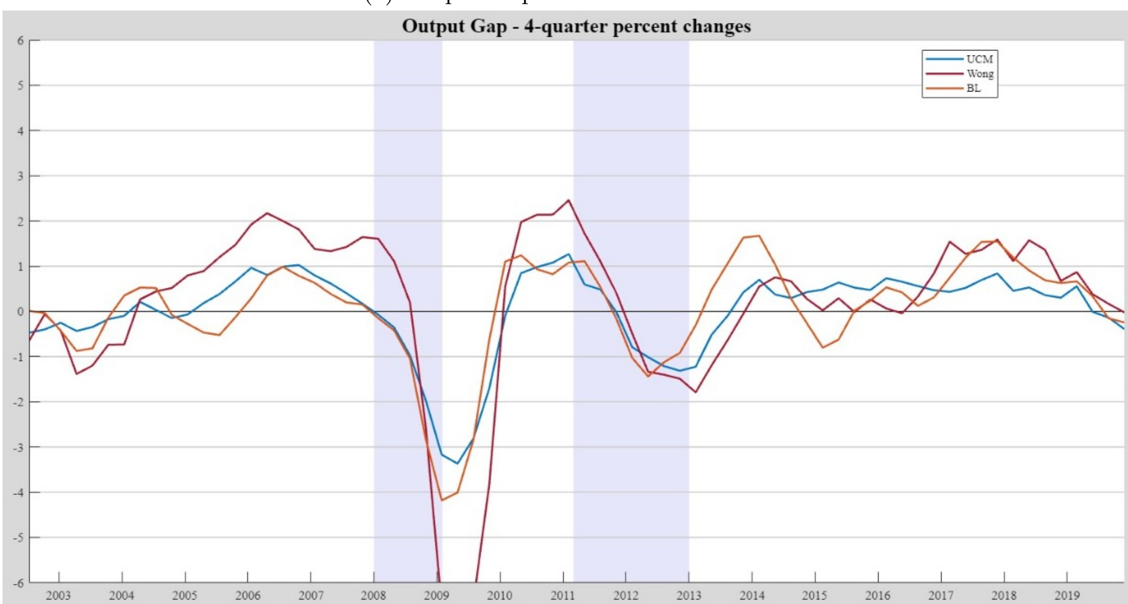
To provide preliminary evidence about the usefulness of an additional factor, we report the results obtained for the BL model using $q = 5$, with figures 8a and 8b reporting respectively the Output Gap in level and the Output Gap in percent changes. Comparing these results with those in figure 1g, which was the closest to the benchmark, we may notice that using 5 factors slightly improves the quality of the predictions. Although estimates are more imprecise at the extremes of the sample period, in the central part of the sample all point estimates are closer to the UCM benchmark, with also a goodness of fit comparable or even better than the one obtained with the MW model. Thus, in general, the results obtained with $q = 5$ seem to be able to solve part of the issues the BL model had in previous analyses.

These results should be an invitation to further analyze both theoretically and empirically which should be the perfect number of factors q when estimating the model in this specific setting. Although the model is not computationally intensive, estimating the results multiple times with a huge basket of possible q_S in order to find the best fit doesn't appear to be the right solution. Indeed, too much room would be left for arbitrariness. Being able to find *a priori* a couple of possible number of factors would instead be the right path to follow. In fact, this would provide a broader understanding of the main drivers of the European economy, laying the ground for an economic interpretation of the factors themselves.

To conclude, as long as these results are comparable or slightly better than the MW estimates, we should add to section 4.1 the following statement: the two models bring similar results in terms of quality of the prediction (using UCM as benchmark), thus the choice between the two has to be done taking into account the trade-off between computation costs and the closeness of the results to UCM estimations. In any case, the results found in this appendix do not cause any threat to the discussion made after point 4.1, which has not been influenced by this analysis.



(a) Output Gap - Level with 5 factors



(b) Output Gap - Percent Changes with 5 factors

Figure 8: BL model results with 5 factors

References

Adrian Alter, Alan Xiaochen Feng, and Nico Valckx. *Understanding the macro-financial effects of household debt: A global perspective.* International Monetary Fund, 2018.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1): 191–221, 2002.

Matteo Barigozzi and Matteo Luciani. Measuring the output gap using large datasets. *Review of Economics and Statistics*, pages 1–15, 2023.

Susanto Basu and John G Fernald. What do we know (and not know) about potential output? *Federal Reserve Bank of St. Louis Review*, 91(July/August 2009), 2009.

Stephen Beveridge and Charles R Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics*, 7(2):151–174, 1981.

Claudio Borio, Piti Disyatat, and Mikael Juselius. Rethinking potential output: Embedding information about the financial cycle. *Oxford Economic Papers*, 69(3):655–677, 2017.

Michael C Burda and Jennifer Hunt. What explains the german labor market miracle in the great recession? Technical report, National bureau of economic research, 2011.

Andrea Carriero, Todd E Clark, and Massimiliano Marcellino. Bayesian vars: specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73, 2015.

CBO. Cbo's method for estimating potential output: An update. 2001.

Alistair Dieppe, Björn van Roye, and Romain Legrand. The bear toolbox. Working Paper Series 1934, European Central Bank, 2016. URL https://EconPapers.repec.org/RePEc:ecb:ecbwps:20161934.

Simon Gilchrist and Benoit Mojon. Credit risk in the euro area. *The Economic Journal*, 128(608):118–158, 2018.

Michael Graff and Jan-Egbert Sturm. The information content of capacity utilization rates for output gap estimates. *CESifo Economic Studies*, 58(1):220–251, 2012.

Alejandro Justiniano and Bruce Preston. Can structural small open-economy models account for the influence of foreign disturbances? *Journal of International Economics*, 81(1):61–74, 2010.

Alejandro Justiniano, Giorgio E Primiceri, and Andrea Tambalotti. Is there a trade-off between inflation and output stabilization? *American Economic Journal: Macroeconomics*, 5(2):1–31, 2013.

Güneş Kamber and Benjamin Wong. Global factors and trend inflation. *Journal of International Economics*, 122: 103265, 2020.

Michael T Kiley. Output gaps. *Journal of Macroeconomics*, 37:1–18, 2013.

Michele Lenza and Giorgio E Primiceri. How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*, 37(4):688–699, 2022.

Massimiliano Marcellino and Alberto Musso. The reliability of real-time estimates of the euro area output gap. *Economic Modelling*, 28(4):1842–1856, 2011.

Michael W McCracken and Serena Ng. Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589, 2016.

James Morley and Benjamin Wong. Estimating and accounting for the output gap with large bayesian vector autoregressions. *Journal of Applied Econometrics*, 35(1):1–18, 2020.

James Morley, Diego Rodríguez-Palenzuela, Yiqiao Sun, and Benjamin Wong. Estimating the euro area output gap using multivariate information and addressing the covid-19 pandemic. *European Economic Review*, 153:104385, 2023.

James C Morley. A state–space approach to calculating the beveridge–nelson decomposition. *Economics Letters*, 75(1):123–127, 2002.

Lee E Ohanian and Andrea Raffo. Aggregate hours worked in oecd countries: New measurement and implications for business cycles. *Journal of Monetary Economics*, 59(1):40–56, 2012.

Arthur Melvin Okun. Potential GNP: Its measurement and significance. *Proceedings of the Business and Economics Statistics Section of the American Statistical Association*, pages 98–104, 1962.

Athanasios Orphanides. The quest for prosperity without inflation. *Journal of monetary Economics*, 50(3):633–663, 2003.

Athanasios Orphanides and Simon van Norden. The unreliability of output-gap estimates in real time. *Review of economics and statistics*, 84(4):569–583, 2002.

James H Stock and Mark W Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002.

James H Stock and Mark W Watson. Disentangling the channels of the 2007-2009 recession. Technical report, National Bureau of Economic Research, 2012.

James H Stock and Mark W Watson. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 415–525. Elsevier, 2016.

Lawrence H Summers. The age of secular stagnation: What it is and what to do about it. *Foreign affairs*, 95(2): 2–9, 2016.

Tao Zha. Block recursion and structural vector autoregressions. *Journal of Econometrics*, 90(2):291–316, 1999.